



Anna T. Waggener, Ph.D.
Institutional Assessment
United States Army War College
Principles of Graphical
Excellence

Best Paper: ALAIR April 5-6, 2001

AIR: June 2-5, 2002, Toronto

FAIR: Feb 21-24, 2003

The Visual Display of Quantitative Information

Leading authority: *Edward R. Tufte*

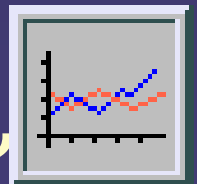
History of Graphical Development

- First geographic maps were drawn on clay tablets.
- 17th Century: combined map skills and statistical skills to construct maps.
- Trade winds and monsoons on a world map.
- Chart patterns of disease.
- Later sophistication showed distribution of 1.3 million galaxies.

***“Graphical
excellence consists
of the efficient
communication of
complex
quantitative ideas.”***

Presentation Topics

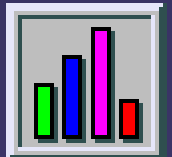
- **Organizing Numerical Data:**
 - **The Ordered Array and Stem-leaf Display**
- **Tabulating and Graphing Numerical Data:**
 - **Frequency Distributions: Tables, Histograms, Polygons**
 - **Cumulative Distributions: Tables, Ogive**



Presentation Topics

(continued)

- **Tabulating and Graphing Univariate Categorical Data:**



- **The Summary Table**

- **Bar and Pie Charts, the Pareto Diagram**

- **Tabulating and Graphing Bivariate Categorical Data:**

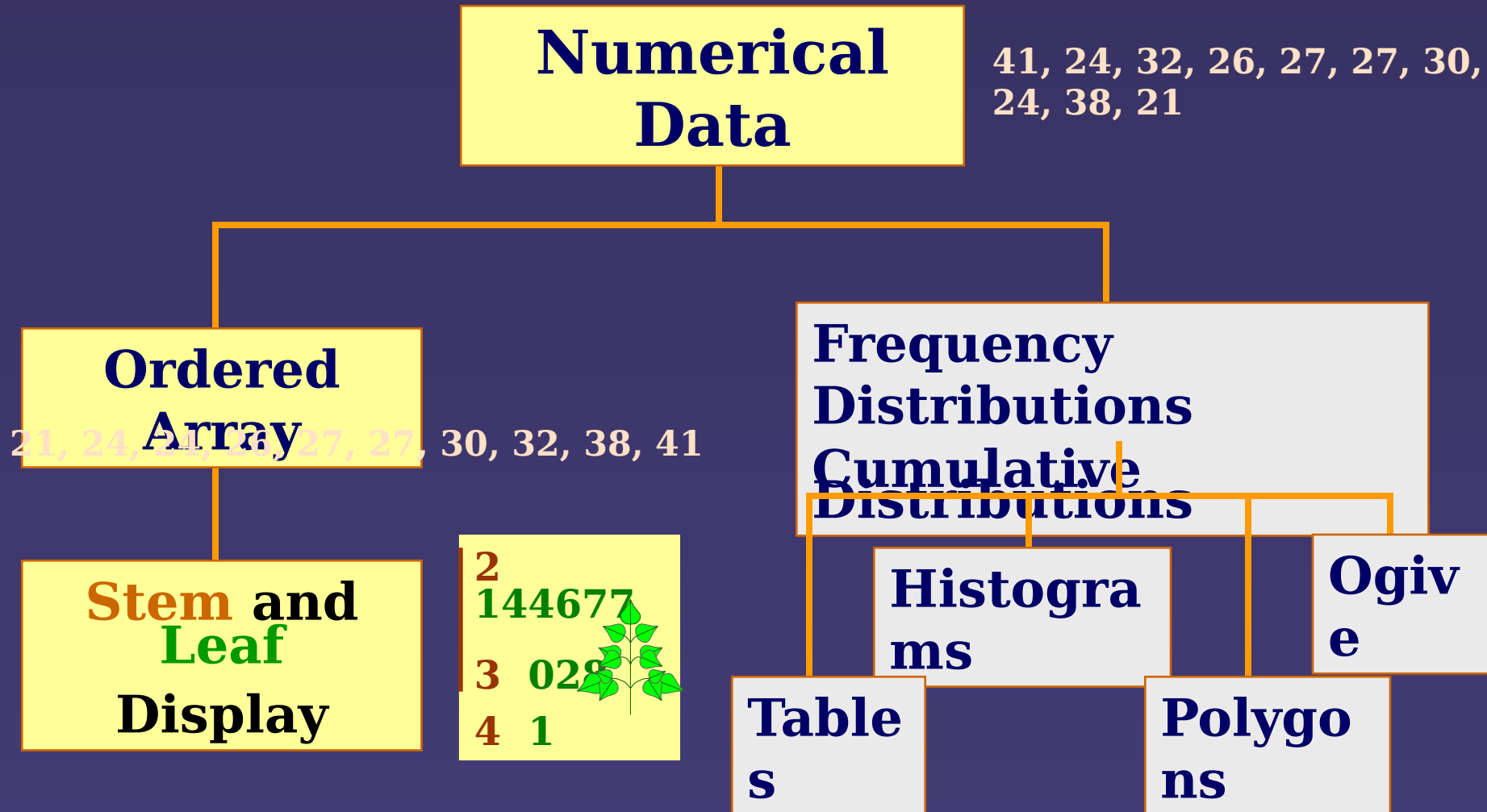
- **Contingency Tables**

- **Side by Side Bar charts**

- **Graphical Excellence and Common Errors in Presenting Data**

***“At their best,
graphics are
instruments for
reasoning about
quantitative
information.”***

Organizing Numerical Data



Organizing Numerical Data:

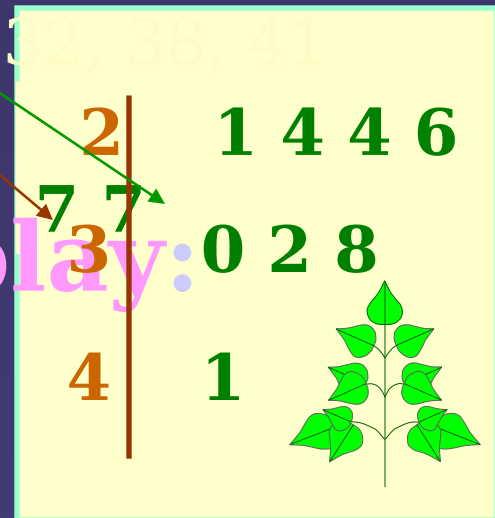
- Data in *Raw* form (as collected):

24, 26, 24, 21, 27, 27, 30, 41, 32, 38

- Data *Ordered* from *Smallest to Largest*:

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Stem and Leaf display:



“Design is choice.”

Tabulating and Graphing Numerical Data

Numerical Data

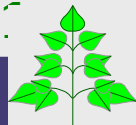
41, 24, 32, 26, 27, 27, 30, 24, 38, 21

Ordered Array

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

Stem and Leaf Display

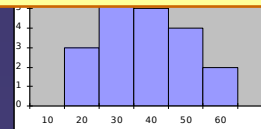
2
144677
3 028
4



Frequency Distributions Cumulative Distributions

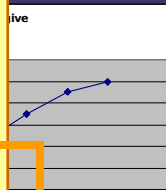
Histograms

Tables



Ogive

Polygons



Tabulating Numerical Data: Frequency Distributions

(continue)

Data in ordered array:

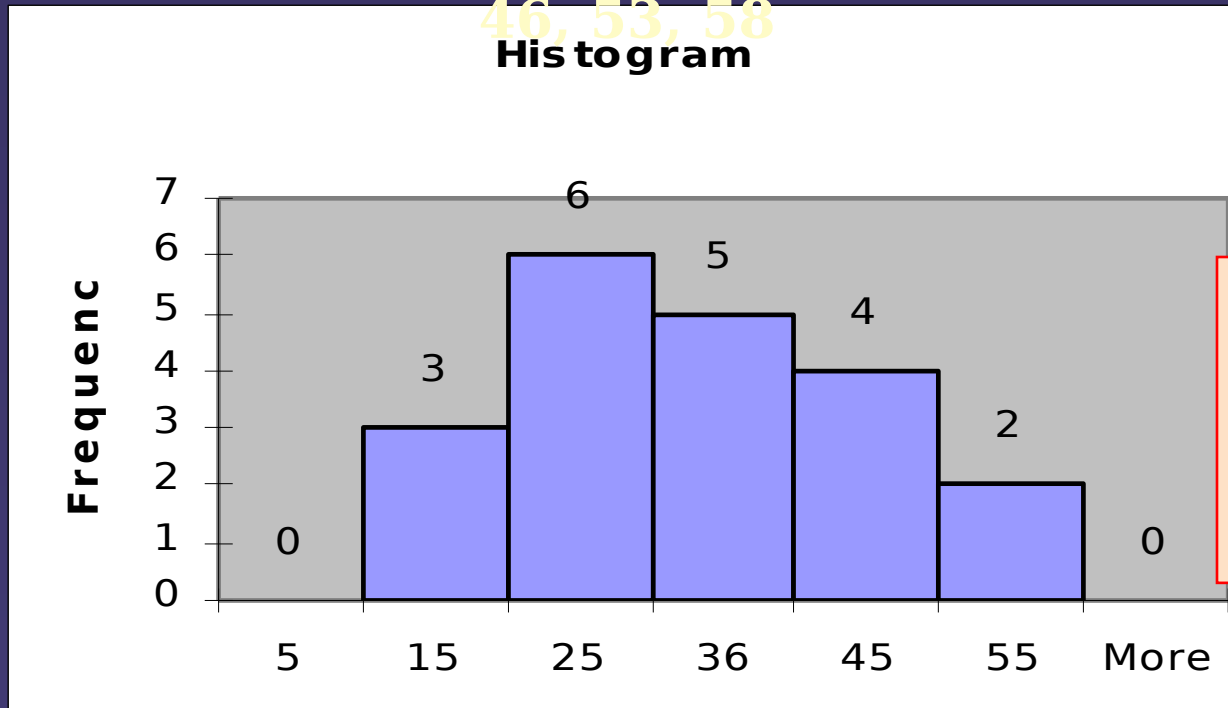
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44,
46, 53, 58

Class	Frequency	Relative Frequency	Percentage
10 but under 20 15	3		.15
20 but under 30 30	6		.30
30 but under 40 25	5		.25
40 but under 50 20	4		.20

Graphing Numerical Data: The Histogram

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44,
46, 53, 58



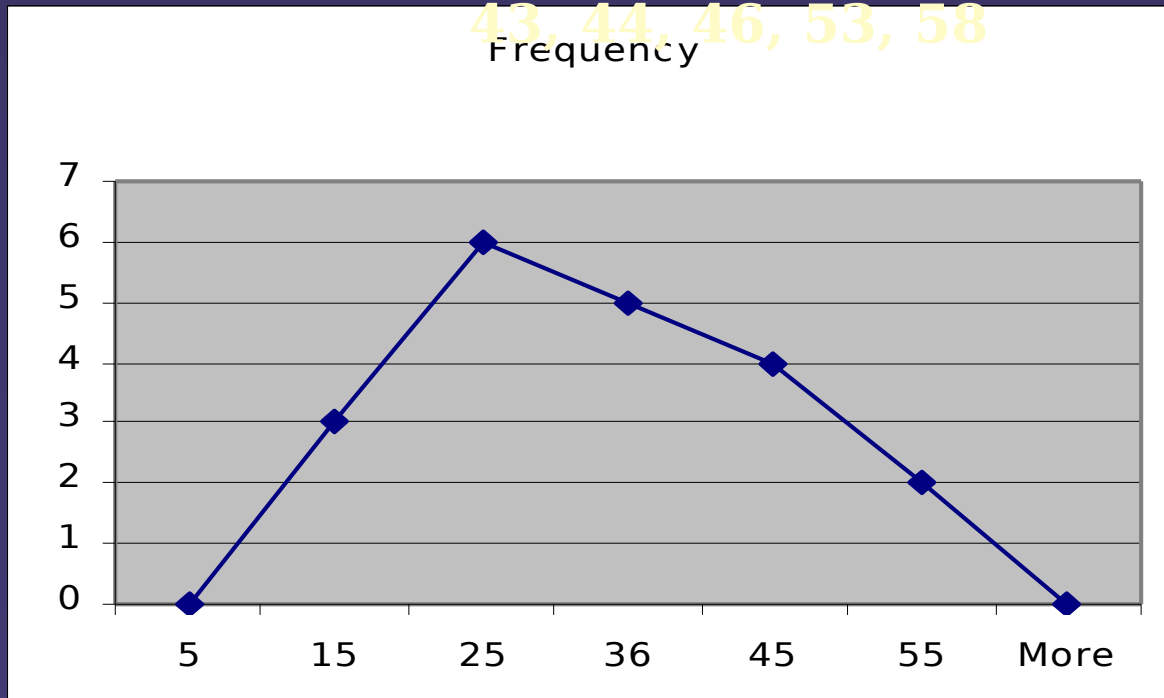
**No
Gaps
Between
n Bars**

**Class
Midpoints**

Graphing Numerical Data: The Frequency Polygon

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41,
43, 44, 46, 53, 58



**Class
Midpoints**

Tabulating Numerical Data: Cumulative Frequency

Data in ordered array:

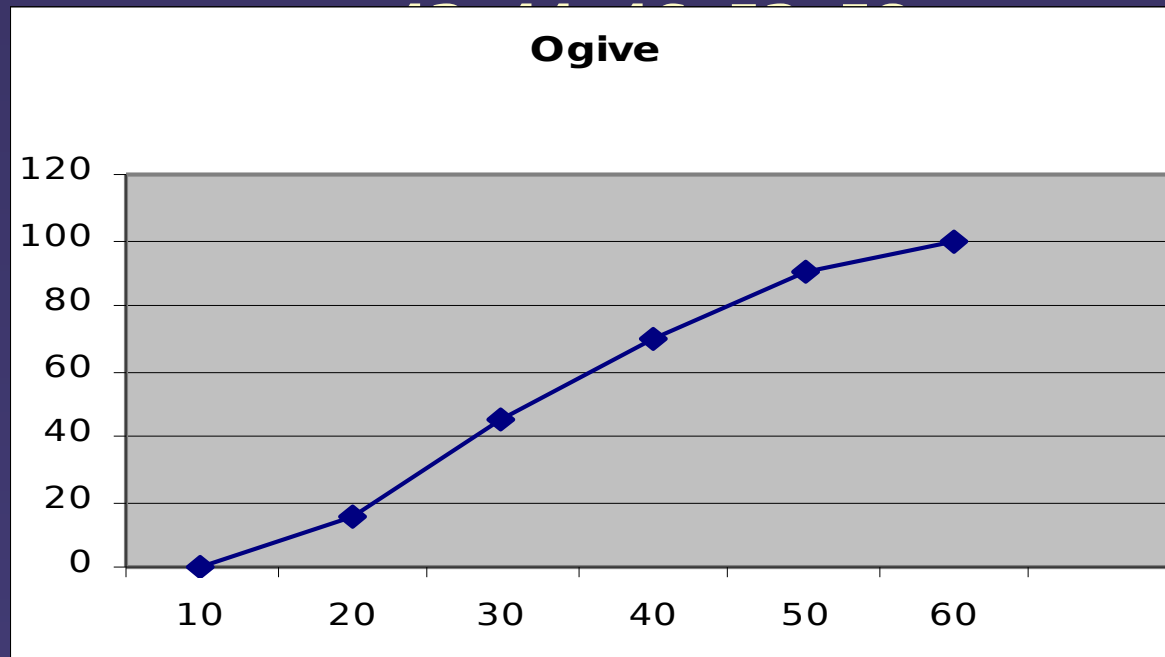
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41,
43, 44, 46, 53, 58

Cumulative Class % Frequency	Cumulative Frequency
10 but under 15	3
20 but under 25	9
30 but under 40	14
40 but under 50	18

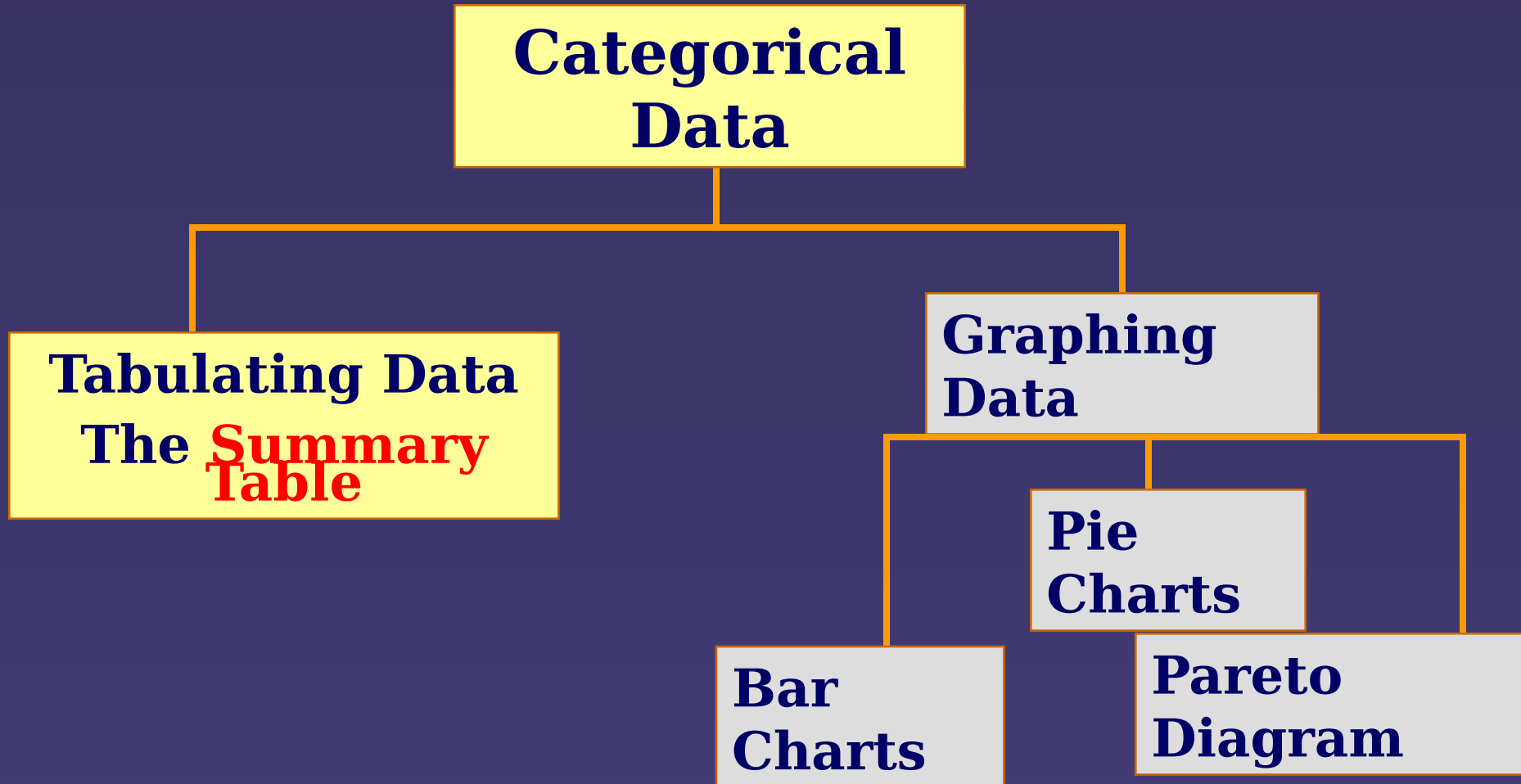
Graphing Numerical Data: The Ogive (Cumulative % Polygon)

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41,



Tabulating and Graphing Categorical Data: Univariate Data



Summary Table

(University Revenues)

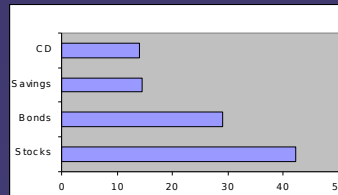
Revenue Category	Amount	
Percentage	(in thousands \$)	
Patient Services	46.5	
Tuition/fees	32	
Appropriations	15.5	14.09
Grants/Contracts	16	
Total	110	

Variables are
Categorical.

Graphing Categorical Data: Univariate Data

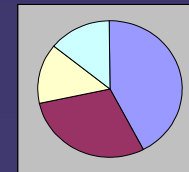
Categorical Data

Tabulating Data
The Summary Table



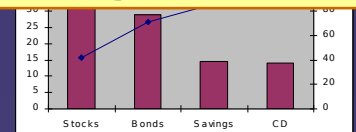
Bar Charts

Graphing Data



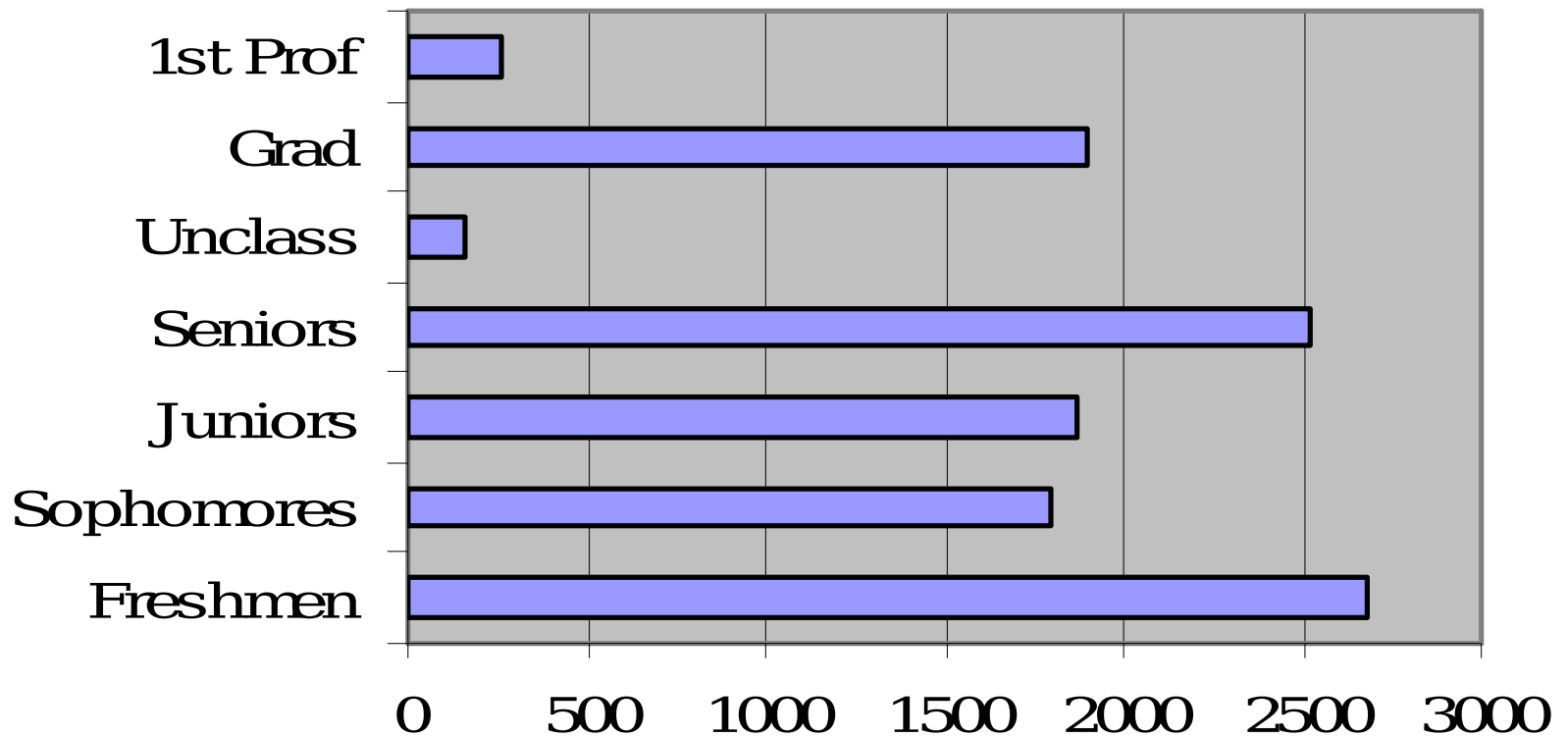
Pie Charts

Pareto Diagram



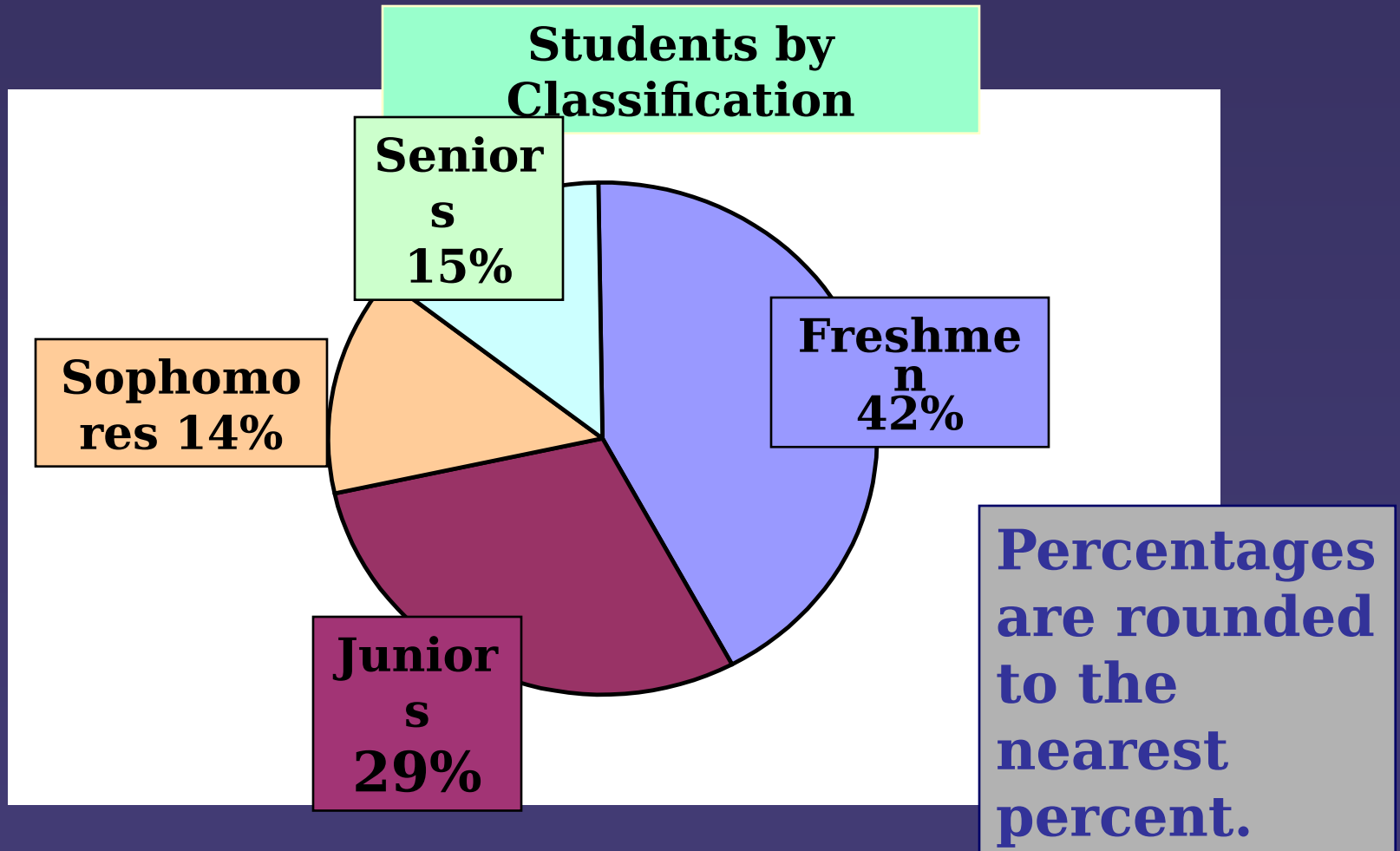
Bar Chart

Enrollment Summary



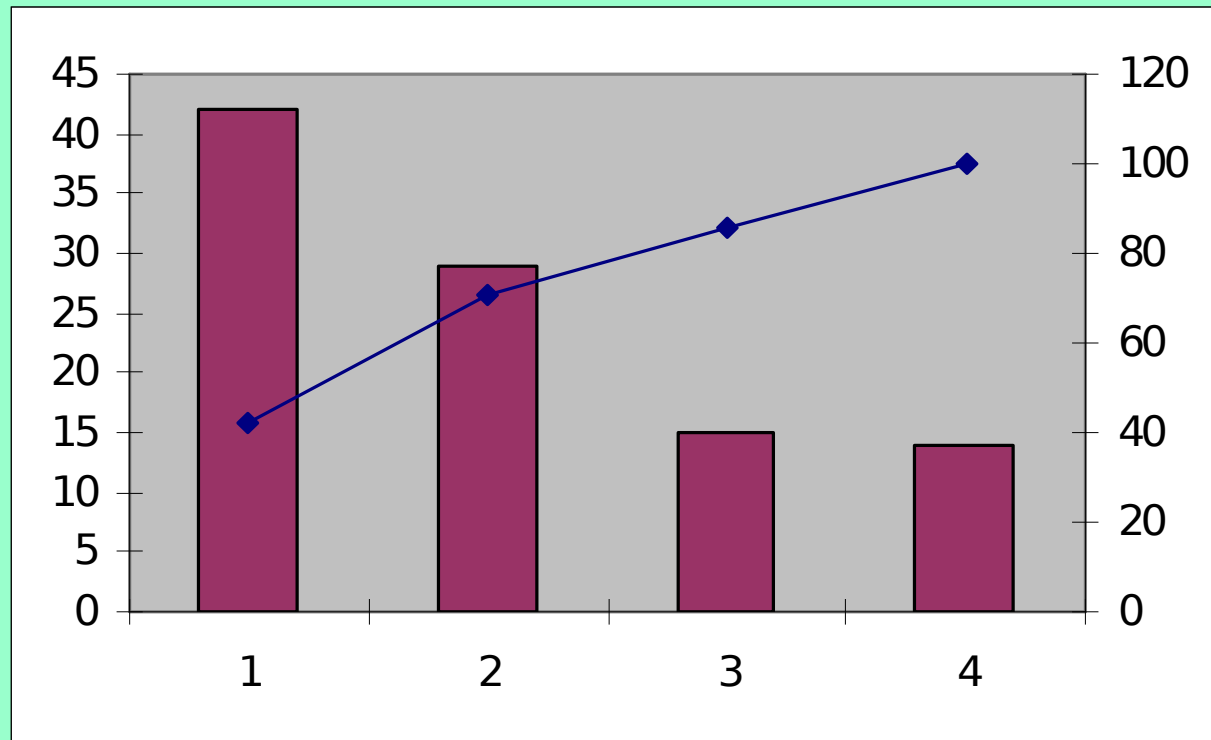
Pie Chart

(for a factbook)



Pareto Diagram

**Axis
for
bar
chart
shows
% in
each
category**



**Axis for
line
graph
shows
cumulative %**

Tabulating and Graphing Bivariate Categorical Data

- **Contingency Tables**
- **Side by Side Charts**

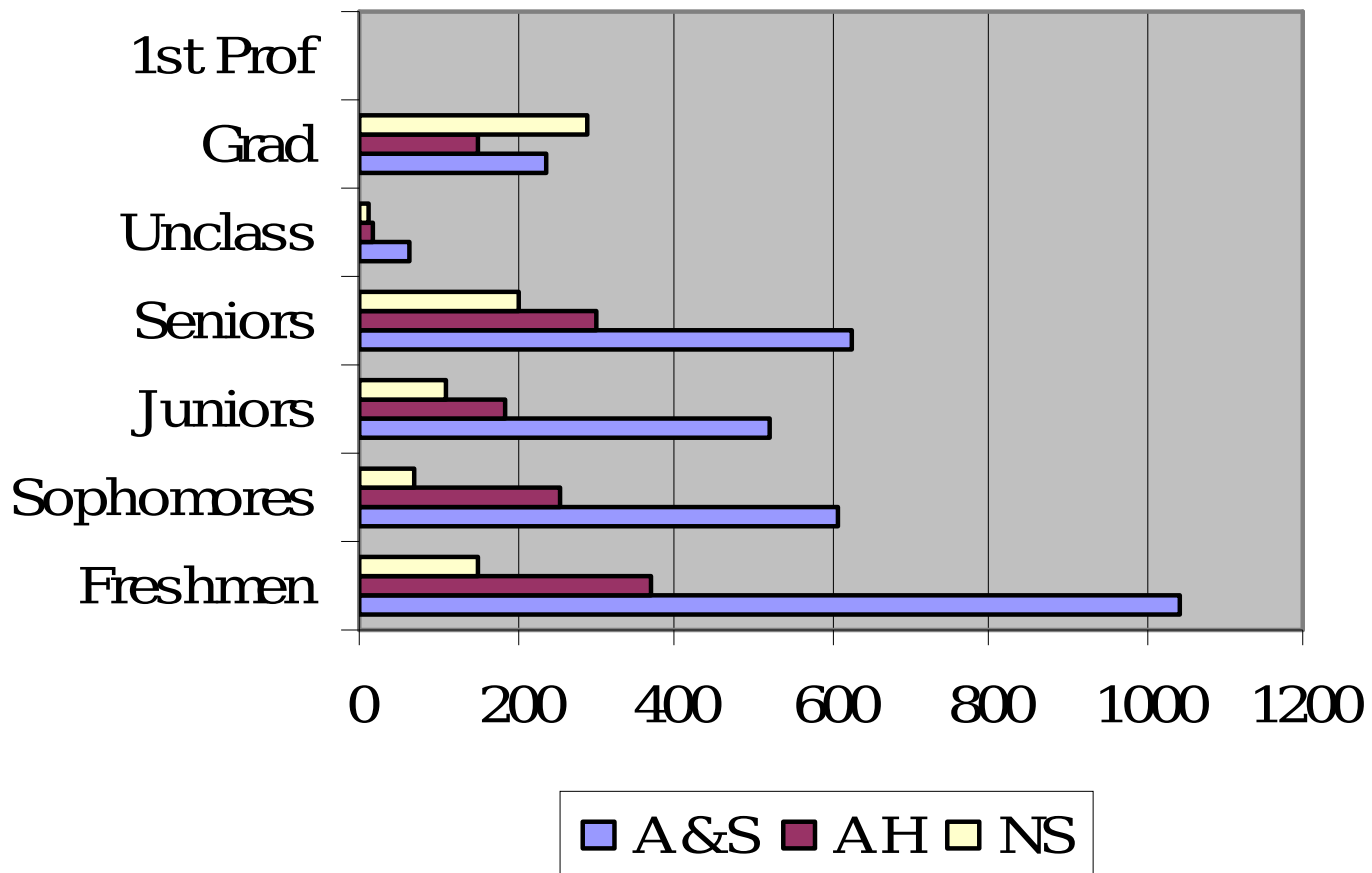
Tabulating Categorical Data: Bivariate Data

Contingency Table: Enrollment by College

Enrollment		A&S		BUS	
NRS		Total		Category	
Freshmen	46	55	27	128	
Sophomores	32	44	19	95	
Juniors	48	15	20	83	
Seniors	16	28	7	51	
Total	322	109	147	66	

Graphing Categorical Data: Bivariate Data

**Side by
Side**



Principles of Graphical Excellence

- Well designed presentation of data that provides:
 - Substance
 - Statistics
 - Design
- Communicates complex ideas with clarity, precision and efficiency
- Gives the largest number of ideas in the most efficient manner
- Almost always involves several dimensions
- Requires telling the truth about the data

Data-Ink Ratio

Data information

Total ink used to print the
graphic

“Much of twentieth-century thinking about statistical graphics has been preoccupied with the question of how some amateurish chart might fool a naive viewer.”

Errors in Presenting Data

- **Using 'chart junk'**
- **No relative basis**
In comparing
data
Batches
- **Compressing the**
Vertical axis
- **No zero point on the**
Vertical axis



'Chart Junk'



Bad Presentation

Minimum Wage



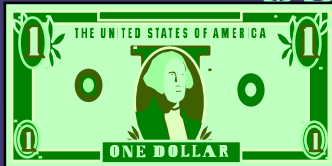
1960:



\$1.00
1970:



\$1.60
1980:



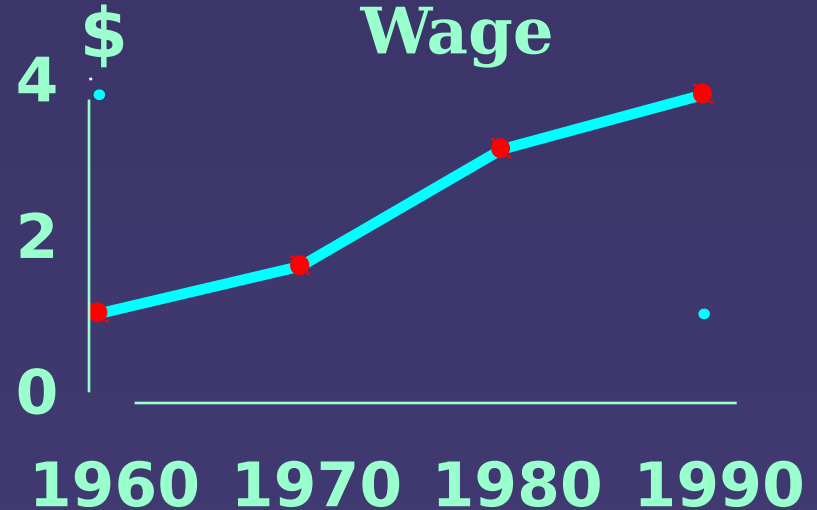
\$3.10

1990:
\$3.80



Good Presentation

Minimum Wage



Lie Factor

Size of effect shown in
graphic

Size of effect in data

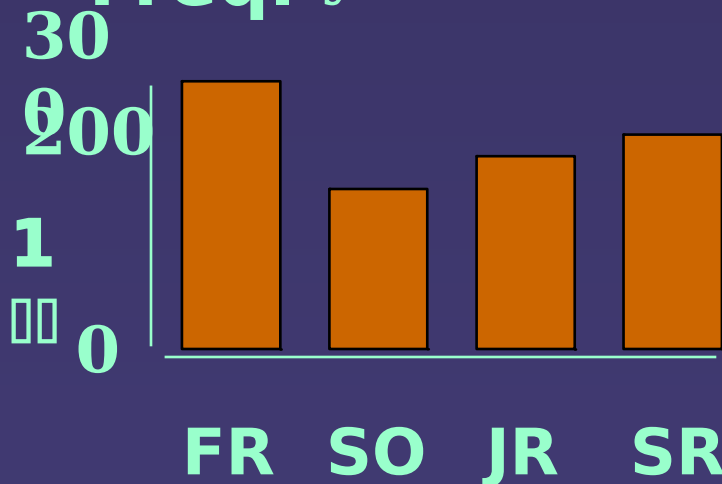
No Relative Basis



Bad

**As received
Presentation**

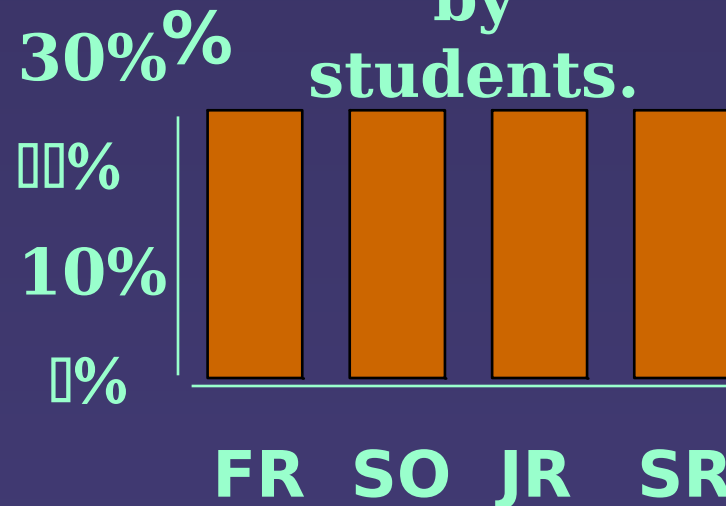
Freq. by students.



Good

**As received
Presentation**

by students.

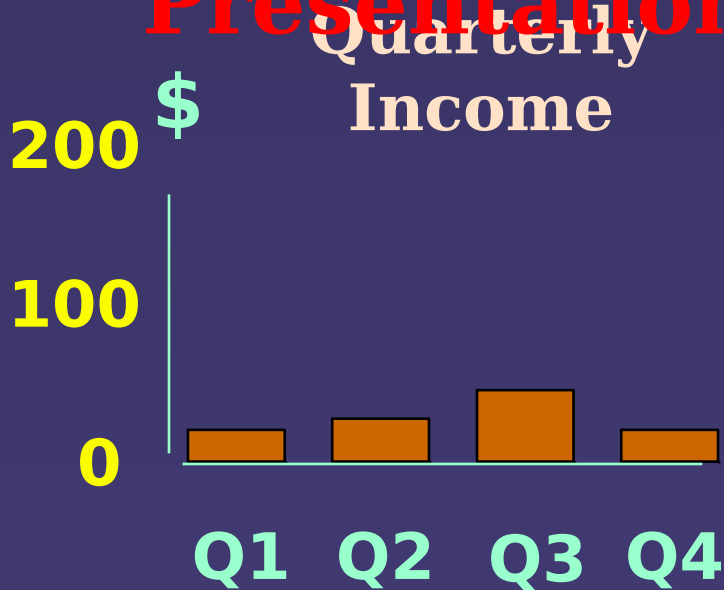


FR = Freshmen, SO = Sophomore, JR = Junior, SR = Senior

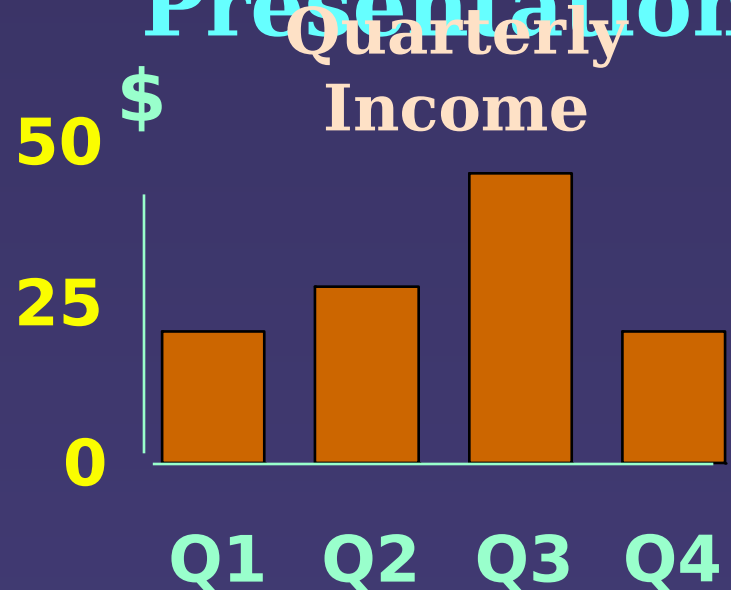
Compressing Vertical Axis



**Bad
Presentation**



**Good
Presentation**



No Zero Point on Vertical Axis



Bad
Presentation
Monthly
Expenses



Good
Presentation
Monthly
Expenses



Graphing the first six months
of sales.

No Zero Point on Vertical Axis



Bad Presentation



Good Presentation



Graphing the first six months
of sales.

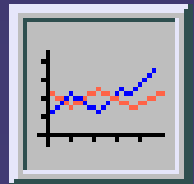
Main defense of the lying graphic....

“Well, at least it was approximately correct, we were just trying to show the general direction of change.”

Presentation

Summary

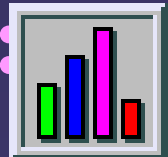
- **Organized Numerical Data:**
 - **The Ordered Array and Stem-leaf Display**
- **Tabulated and Graphed Numerical Data**
 - **Frequency Distributions: Tables, Histograms, Polygons**
 - **Cumulative Distributions: Tables, Ogive**



Presentation

Summary *(continued)*

- Tabulated and Graphed
Univariate Categorical Data:



- The Summary Table
- Bar and Pie Charts, the Pareto diagram

- Tabulated and Graphed
Bivariate Categorical Data:

- Contingency Tables
- Side by Side charts
- Discussed Graphical Excellence and Common Errors

*“There remain, however,
many other consideration
in the design of
statistical graphics - not
only of efficiency, but
also of complexity,
structure, density, and
even beauty.”*

***“Without data, it is
anyone’s
opinion.”***